

Ширабайкин Иван Анатольевич
магистрант МТУСИ, Москва, Россия,

СОВРЕМЕННЫЕ МЕТОДЫ МЕЖКАМЕРНОГО ТРЕКИНГА НА ОСНОВЕ КЛАСТЕРИЗАЦИИ БИОМЕТРИЧЕСКИХ ПРИЗНАКОВ ЛИЦА

Аннотация

В статье рассматриваются методы межкамерного трекинга лиц, основанные на кластеризации биометрических эмбеддингов и учёте пространственно-временных ограничений. Проанализированы подходы, использующие скользящие временные окна, глобальную кластеризацию траекторий и представление идентичностей в виде центров кластеров признаков. Показано, что совместное использование признаков внешнего вида и временных моделей позволяет повысить устойчивость идентификации и снизить количество ошибок ассоциации в многокамерных системах видеонаблюдения. Полученные результаты могут быть использованы при разработке и модернизации практических систем межкамерного трекинга.

Ключевые слова: *межкамерный трекинг, МТМСТ, лицевые эмбеддинги, биометрия, повторное распознавание личности, кластеризация, центры кластеров, временное окно, пространственно-временные ограничения.*

Abstract

The paper considers methods for multi-camera face tracking based on clustering of biometric face embeddings and the use of spatio-temporal constraints. Approaches employing sliding time windows, global trajectory clustering, and identity representation via cluster centers are analyzed. It is shown that the joint use of appearance features and temporal models improves identification stability and reduces association errors in multi-camera surveillance systems. The obtained

results can be applied in the development and modernization of practical multi-camera tracking systems.

Keywords: *multi-camera tracking, MTMCT, face embeddings, biometrics, re-identification, clustering, cluster centres, temporal window, spatio-temporal constraints.*

Введение

Межкамерный трекинг (*Multi-Camera Tracking, MTMCT*) направлен на сохранение идентичности субъектов при их перемещении через зоны наблюдения нескольких разнесённых камер, что является важной задачей компьютерного зрения в системах видеонаблюдения, аналитике поведения и робототехнике. Особенность межкамерного сценария в том, что камеры могут находиться далеко друг от друга, их поля зрения не перекрываются, что приводит к длительным периодам исчезновения объекта из поля зрения (окклюзиям) и существенным различиям в ракурсах и освещении между разными камерами. Кроме того, число объектов обычно заранее неизвестно, а объём обрабатываемых данных достаточно велик [1, 2].

Современные системы межкамерного трекинга, подобно однокамерным МОТ-системам, как правило, следуют парадигме «обнаружение – сопоставление» (*tracking-by-detection*). Сперва выполняется детекция и отслеживание объектов отдельно в каждом видеоканале, а затем результаты (траектории) сопоставляются между камерами [1, 2]. Ключевым шагом межкамерного трекинга является ассоциация траекторий одного и того же объекта из разных камер, которая использует комбинацию признаков внешнего вида объекта и пространственно-временных ограничений.

В последние годы особенно успешно стало применение алгоритмов повторного распознавания личности (*person re-identification, Re-ID*) для оценки сходства появлений объекта в разных камерах. Глубокие нейросетевые модели *Re-ID* способны извлекать дискриминативные признаки внешности, позволяющие надёжно сопоставлять людей между разными камерами [4]. В

контексте межкамерного трекинга лиц в качестве признакового описания используются векторы эмбедингов, отражающие биометрические характеристики лица, что позволяет строить устойчивые идентичности в пространстве признаков и далее применять методы кластеризации для их объединения. Однако одного совпадения по внешнему виду недостаточно, из-за потенциальной схожести разных людей и изменчивости условий съёмки необходим учёт временных и геометрических факторов [2, 4].

В настоящей работе обобщаются наиболее эффективные подходы к межкамерному трекингу, включая новейшие методы, основанные на использовании скользящих временных окон, априорных временных ограничений и кластерных центров траекторий и эмбедингов. Особое внимание уделяется решениям, в которых кластеризация биометрических признаков (прежде всего лицевых эмбедингов) интегрируется с пространственно-временной моделью переходов между камерами.

Временные окна и временные ограничения

Одним из эффективных подходов к межкамерной ассоциации является ограничение задачи отслеживания скользящим временным окном. Вместо одновременного глобального сопоставления всех наблюдений за весь промежуток времени алгоритм разбивает задачу на более короткие интервалы, что значительно снижает вычислительную сложность и уменьшает вероятность ошибочных соответствий [1]. Ассоциация траекторий выполняется итеративно в пределах данных окон и при отслеживании, как правило, учитываются лишь недавние наблюдения в ограниченном интервале времени, после чего окно последовательно сдвигается по мере поступления новых детекций. Схематически принцип скользящего временного окна при межкамерной ассоциации траекторий представлен на рисунке 1.

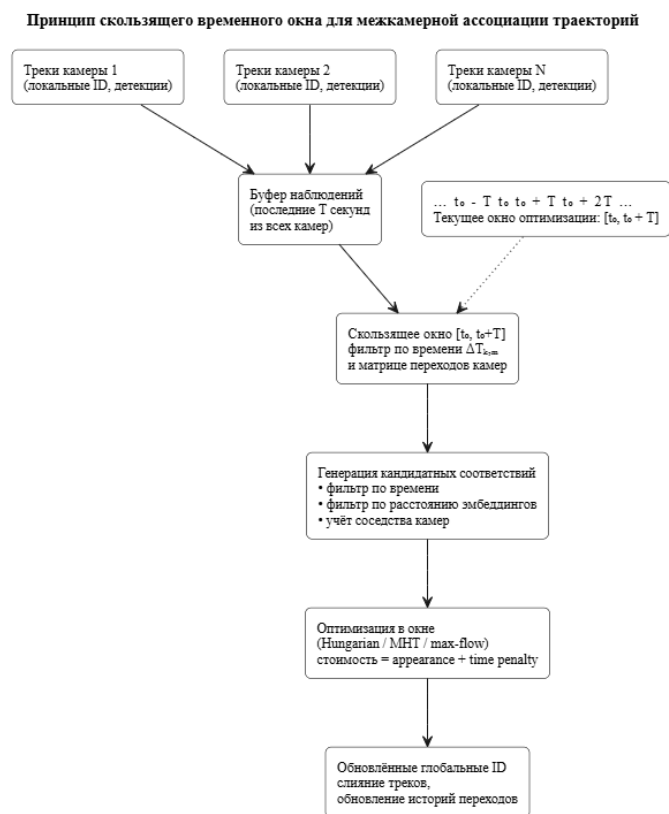


Рисунок 1. Принцип скользящего временного окна для межкамерной ассоциации траекторий

Yoon и соавт. предложили использовать алгоритм множественных гипотез (*Multiple Hypothesis Tracking*) для межкамерного трекинга в реальном времени, формируя дерево гипотез траекторий в движущемся временном окне [1]. Данный подход позволяет учитывать несколько альтернативных путей соответствия при неопределённости (перекрытия, частичные окклюзии), при этом ограничение по окну времени обеспечивает управляемый размер задачи ассоциации и делает метод пригодным для работы в реальном времени.

В условиях офлайн решения (когда доступна вся видеопоследовательность целиком) временные ограничения также остаются важным элементом. Глобальные оптимизационные методы могут достигать более высокой точности ассоциации, учитывая все данные сразу, однако они также накладывают априорные ограничения на допустимое время перемещения объекта между камерами [2]. Известно, что реальное время перехода объекта из области обзора одной камеры в область другой ограничено

и поддаётся моделированию. Для большинства сцен можно построить карту переходов «камера – камера», задавая допустимые интервалы времени для каждой пары камер.

Так, *Maksai* и коллеги рассматривают глобально согласованный трекинг множества объектов, где при оптимизации траекторий учитываются не только локальные соответствия, но и непротиворечивость общей временной структуры, что задаёт скрытые временные окна для допустимых ассоциаций [2]. *Ristani* и *Tomasi* вводят жёсткие временные рамки, исключая нереалистично длинные кандидатные соответствия траекторий между камерами [4]. Дополнительно предлагается механизм затухания корреляции по мере увеличения временного разрыва. Если между двумя появлениями (детекциями), предположительно принадлежащими одному человеку, проходит значительный интервал времени, связь между ними постепенно ослабляется, и для их объединения требуется наличие промежуточной последовательности наблюдений [4].

Кластеризация и центры кластеров для идентификации

Для решения задачи глобальной межкамерной ассоциации широко применяются методы кластеризации траекторий на основе сходства их признаков. В простейшем случае траектории, принадлежащие одной личности, должны образовывать плотные кластеры в пространстве эмбедингов, тогда как траектории разных людей разнесены. Данное утверждение справедливо как для признаков всего тела (*person Re-ID*), так и для биометрических лицевых эмбедингов. Концептуальная иллюстрация кластеризации лицевых эмбедингов с формированием центров кластеров показана на рисунке 2.

Схема кластеризации лицевых эмбеддингов и центров кластеров

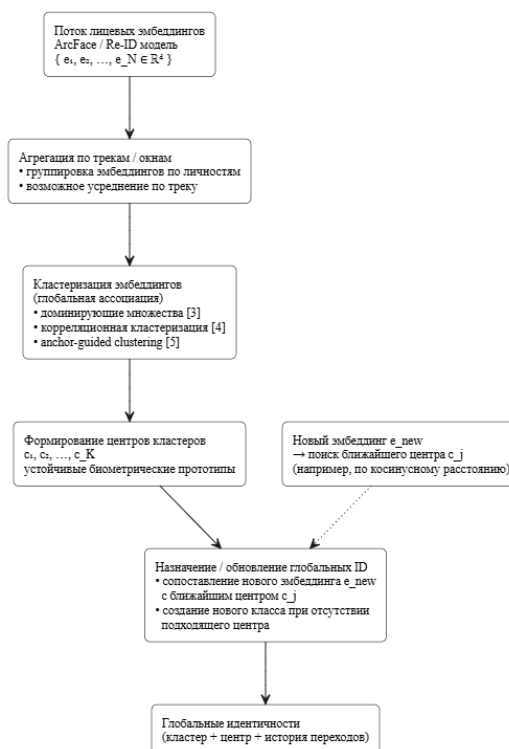


Рисунок 2. Кластеризация эмбеддингов лица в пространстве признаков

В работе *Tesfaye* и др. предложена кластеризация на основе доминирующих множеств (*Constrained Dominant Sets*) для объединения треков одного человека из разных камер [3]. Метод формулирует задачу как поиск устойчивых подмножеств в графе сходства между траекториями, при этом учитываются ограничения на совместимость по времени и пространству, что позволяет выполнять глобальную ассоциацию без жёсткого задания числа кластеров и демонстрирует высокую устойчивость к шумовым наблюдениям [3].

Другие подходы формулируют задачу как кластеризацию графа, где узлами служат траектории или отдельные детекции, а рёбрами – меры сходства между ними. Так, применение корреляционной кластеризации (*maximizing agreement clustering*) ко всем парам наблюдений позволяет объединять идентичности с учётом всех попарных связей и оптимизировать соответствия глобально [4]. Методы, учитывающие все парные сравнения между наблюдениями (не только ближайшие по времени), действительно показывают более высокую точность, хотя задача при этом становится вычислительно

труднее [4]. В практике *Ristani* и *Tomasi* корреляционная кластеризация используется для глобального объединения треков, что позволяет получить точное и при этом достаточно эффективное решение задачи идентификации на нескольких камерах [4].

Современное развитие перечисленных идей представлено в новейших алгоритмах, где кластеризация интегрируется с этапом обучения признаков и пространственно-временной пост-обработкой. *Huang* и коллеги предложили подход *anchor-guided clustering*, в котором выполняется иерархическая кластеризация признаков всех предварительно полученных траекторий для выделения прототипных центров кластеров, представляющих каждого уникального человека [5]. Таким образом каждый центр кластера аккумулирует разнообразные визуальные признаки одной личности, полученные под разными углами зрения и при различных условиях съёмки [5].

После формирования кластерных центров выполняется итеративное уточнение и переназначение идентификаторов с учётом пространственно-временной согласованности (например, сопоставляя траектории по положению на общей карте сцены) для устранения разрывов и ошибок идентификации [5]. Предложенный подход продемонстрировал высокую эффективность на соревновании *CVPR AI City Challenge*, где достиг лучшего показателя *IDF1* среди представленных решений [5]. Таким образом, кластеризация траекторий с опорой на их центроиды (центры кластеров) в сочетании с проверкой пространственно-временного ограничения согласованности позволяет значительно повысить точность межкамерного отслеживания объектов.

Отдельный класс исследований фокусируется на межкамерном трекинге именно по лицевой биометрии. В качестве признаков используются эмбединги лица, а глобальная ассоциация строится вокруг центров кластеров эмбедингов. Концептуально данные методы следуют тем же принципам, что и общие схемы *MTMST* – накопление разнообразных представлений лица в виде кластерных центров, применение корреляционной или иерархической

кластеризации и явный учёт временных ограничений перехода между камерами. В результате удаётся строить прототипы «глобальной личности» и сопоставлять новые наблюдения не с отдельными детекциями, а с устойчивыми центрами кластеров лиц, что уменьшает вероятность ошибок при меняющемся ракурсе и частичных окклюзиях.

Выводы

Современные системы межкамерного трекинга демонстрируют наилучшие результаты при совместном использовании методов сравнения внешнего вида и строгих пространственно-временных ограничений. Применение скользящих временных окон при ассоциации траекторий позволяет существенно сократить размер задачи и уменьшить число ошибок сопоставления без потери полноты отслеживания [1, 2]. Кластерные методы глобальной ассоциации, включая подходы на основе доминирующих множеств, корреляционную кластеризацию и *anchor-guided clustering*, обеспечивают непрерывность и высокую точность идентификации объектов при переходе между камерами [3–5]. Обобщённая архитектура межкамерного трекинга, объединяющая локальный однокамерный трекер, модуль биометрических эмбедингов и глобальную ассоциацию на основе временных окон и кластеров, представлена на рисунке 3.

Обобщённая архитектура системы межкамерного трекинга лиц (компактная схема)



Рисунок 3. Обобщённая архитектура системы межкамерного трекинга

Для межкамерного трекинга лиц особую роль играет кластеризация биометрических эмбедингов лица: формирование центров кластеров, аккумулирующих разнообразные ракурсы и условия наблюдения, позволяет существенно повысить устойчивость идентификаторов. Исследования подтверждают эффективность объединения временных ограничений и

кластеризации идентичностей, демонстрируя заметный прирост метрик отслеживания по сравнению с более ранними решениями, не учитывавшими данные факторы [3, 4, 5].

Целью настоящей работы являлась систематизация современных подходов к межкамерному трекингу лиц и качественное обоснование их практической применимости. Сформулированные выводы могут быть использованы разработчиками промышленных систем видеонаблюдения и исследователями, работающими над задачами *MTMCT*, при проектировании модулей межкамерной ассоциации, ориентированных на использование биометрических лицевых признаков.

Список литературы

1. Yoon K. et al. Multiple hypothesis tracking algorithm for multi-target multi-camera tracking with disjoint views // *IET Image Processing*. 2018. 12(7). 1175–1184.
2. Maksai A. et al. Non-Markovian globally consistent multi-object tracking // *Proc. ICCV*. 2017. P. 2563–2573.
3. Tesfaye Y. T. et al. Multi-target tracking in multiple non-overlapping cameras using fast-constrained dominant sets // *Int. J. Comput. Vis*. 2019. 127(9). 1303–1320.
4. Ristani E., Tomasi C. Features for multi-target multi-camera tracking and re-identification // *Proc. CVPR*. 2018. P. 6036–6046.
5. Huang H.-W. et al. Enhancing multi-camera people tracking with anchor-guided clustering and spatio-temporal consistency ID re-assignment // *Proc. CVPR Workshops (AI City Challenge)*. 2023. P. 5239–5248.